

Learning radiological appearances of diseases: Does comparison help?

Citation for published version (APA):

Kok, E. M., de Bruin, A. B. H., Robben, S. C. F., & van Merriënboer, J. J. G. (2013). Learning radiological appearances of diseases: Does comparison help? *Learning and Instruction*, 23, 90-97.
<https://doi.org/10.1016/j.learninstruc.2012.07.004>

Document status and date:

Published: 01/02/2013

DOI:

[10.1016/j.learninstruc.2012.07.004](https://doi.org/10.1016/j.learninstruc.2012.07.004)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Learning radiological appearances of diseases: Does comparison help?

Ellen M. Kok^{a,*}, Anique B.H. de Bruin^{a,b}, Simon G.F. Robben^c, Jeroen J.G. van Merriënboer^a

^a Maastricht University, School of Health Professions Education, Department of Educational Development and Research, P.O. Box 616, 6200 MD Maastricht, The Netherlands

^b Institute of Psychology, Erasmus University Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands

^c Maastricht University Medical Center, Department of Radiology, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands

ARTICLE INFO

Article history:

Received 16 September 2011

Received in revised form

10 July 2012

Accepted 13 July 2012

Keywords:

Radiology

Visual diagnosis

Education

Comparison

ABSTRACT

Comparison learning is a promising approach for learning complex real-life visual tasks. When medical students study radiological appearances of diseases, comparison of images showing diseases with images showing no abnormalities could help them learn to discriminate relevant, disease-related information. Medical students studied 12 diseases on chest x-ray images. They were randomly assigned to a group ($n = 31$) that compared radiographs of diseases with normal images and a group ($n = 30$) that only studied radiographs of diseases. On a visual diagnosis test, students who compared with normal images during study were better able to diagnose focal diseases (i.e., lesions at one location) than students who could not compare, but for the diagnosis of diffuse diseases (i.e., involving both lungs) there was no significant difference between groups. Results show that comparison with normal images made it easier to discriminate relevant information for focal diseases.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Learning by comparison is a commonly studied topic in educational psychology, and one of its leading researchers, Dedre Gentner, even argues that comparison learning is one of the key processes by which people learn (2010). Learning by comparison is broadly found to be very effective in the context of, for example, category learning (Andrews, Livingston, & Kurtz, 2011), schema acquisition (Gick & Paterson, 1992), and conceptual change (Gadgil, Nokes-Malach, & Chi, 2012). Much research on comparison learning is conducted using artificial tasks. Although it is commonly studied in real-life studies of learning mathematics (see Rittle-Johnson & Star, 2011, for an overview), comparison learning has hardly been applied in other real-life tasks such as complex visual tasks (but see Ark, Brooks, & Eva, 2007, and Hatala, Brooks, & Norman, 2003, for examples of comparison for learning the interpretation of ECGs). Complex visual tasks such as classification in biology, interpretation of weather maps, and visual diagnosis in medicine seem particularly fit for comparison learning. A key aspect of expertise in complex visual tasks is the ability to discriminate (Kellman & Garrigan, 2009). Comparison of contrasting exemplars (i.e., stimuli belonging to different categories) is an excellent way to learn discrimination (Andrews et al., 2011; Hammer, Bar-Hillel, Hertz, Weinshall, & Hochstein, 2008).

We investigated the effect of comparison by studying its effect on learning a prototypical real-life complex visual task: diagnosing conventional chest radiographs (x-ray images of the chest). Furthermore, we focus on different effects of comparison learning for different types of images.

1.1. Comparison learning for real-life complex visual tasks

Comparison of contrasting exemplars (two or more examples that belong to different categories) helps to learn discrimination (Hammer et al., 2008). The ability to discriminate is a key aspect of expertise in visual skills (Kellman & Garrigan, 2009). For example, while a novice bird-watcher might be able to discriminate between a sparrow and an owl (making little distinction between different types of owls), an experienced bird-watcher has obtained the ability to discriminate more specifically, for example between a great gray owl and a northern hawk owl (Tanaka, Curran, & Sheinberg, 2005). An example of discrimination in chest radiographs is shown in Fig. 1: The lower outer corners of the lungs (sinuses) normally curve downward. In order to discriminate pleural effusion from a normal image, it is important to realize that in pleural effusion the liquid is curving upwards at the sides of the lungs. The direction of curving discriminates normality from pleural effusion.

Resulting from the ability to discriminate more dimensions in the stimulus array is the ability of experts to discriminate relevant from irrelevant information (Gibson, 1969). Experts in a domain are more likely to ignore task-irrelevant and redundant information

* Corresponding author. Tel.: +31(0)433885761; fax: +31(0)43 388 5779.
E-mail address: e.kok@maastrichtuniversity.nl (E.M. Kok).

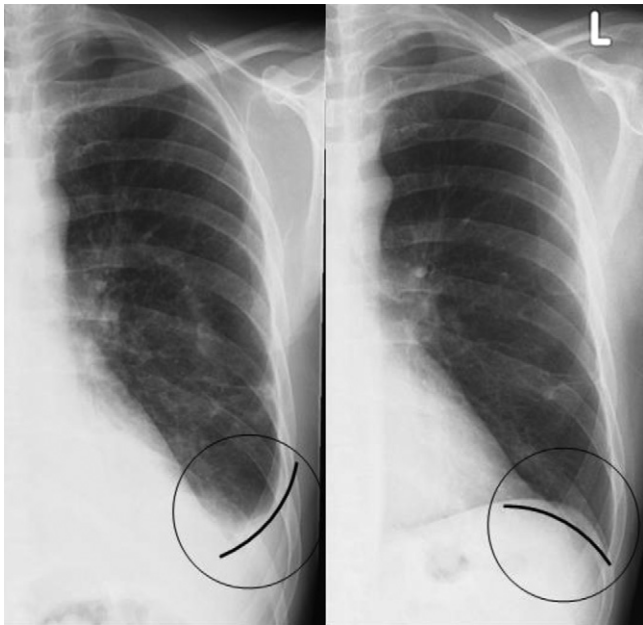


Fig. 1. Subtle features of a disease. Left: pleural effusion, left sinus is curving upward, right: normal image, left sinus is curving downward.

and focus their attention specifically on task-relevant information, leading to an optimized amount of processed information (Gegenfurtner, Lehtinen, & Säljö, 2011). For example, in a classic study on chick sexing, experts knew exactly which information was relevant for determining the sex of day-old chicks, leading to the ability to sex 98% of the chicks correct at a rate of 1000 chicks per hour, while novices did not know which information to use for discrimination between male and female chicks (Biederman & Shiffrar, 1987). In the medical domain, Balslev and colleagues found that pediatricians looked more often to task-relevant body parts (showing abnormal movements) rather than task-irrelevant body parts (showing no abnormal movements) in videos of infants having epileptic seizures, compared to medical students (Balslev et al., 2011).

The phenomenon that novices in a domain have problems discriminating relevant from irrelevant information in complex visual displays is commonly found in real-life complex visual tasks (e.g., Jarodzka, Scheiter, Gerjets, & van Gog, 2010; Lowe, 1999; Wood, 1999). In those kinds of tasks, a lot of information is present and not all information is task-relevant. Novices are more likely to attend to information based on conspicuity than on relevance, even if this conspicuous information is not relevant (Lowe, 1999). Finding the location of relevant information and ignoring conspicuous yet irrelevant information in a visual display is crucial, though, because if students are not able to attend to the relevant information in a complex visual display, they will naturally not be able to learn this information (Boucheix & Lowe, 2010).

Diagnosing radiological images is a typical example of a task in which the discrimination of relevant from irrelevant information is problematic for novices (Wood, 1999). Radiological images, such as conventional chest radiographs (x-ray images of the chest), contain a wealth of information that needs to be interpreted for visual diagnosis. A lot of this information is not related to diseases (Mettler, 2005). For example, on conventional chest radiographs, women's breasts make the tissue behind the breasts appear whiter, while nipples' shadows may look like tumors. Real tumors, on the other hand, might be masked by adjacent ribs (Samei, Flynn, Peterson, & Eyler, 2003). Furthermore, a radiograph is

a two-dimensional representation of a three-dimensional object, so there is also overprojection, and the size of an organ on the radiograph depends on the distance to the detector (Mettler, 2005). These phenomena can make normal tissue appear suspicious and mask abnormalities, making the discrimination of relevant from irrelevant information a difficult but necessary task.

While students are studying radiological images to learn the appearance of diseases, they have to discriminate between information that is relevant for diagnosing a disease and information that is not disease-related. Relevant information for diagnosing a disease has to be incorporated in the mental representation of that disease, while information that is not disease-related should be left out. For example, the shape of the chest is not relevant for the diagnosis of a tumor and should not be incorporated in the mental representation of the appearance of a tumor. The quality of the mental representation of a disease influences later visual diagnostic performance (c.f. Lowe, 2005).

Comparison of contrasting images can help students to isolate relevant but less conspicuous information (Gentner & Gunn, 2001). For example, when comparing two pictures of offices, one that has a computer and one that has no computer, the computer is easy to find and likely to attract attention. However, if you would view only the picture of the office with the computer, the computer would not draw special attention (Gentner & Gunn, 2001). According to structural alignment theory (Markman & Gentner, 1997), during comparison of stimuli, features and relations within one stimulus are systematically matched to features and relations in the other stimulus (i.e., aligned). Differences between two stimuli become more salient as a result of this matching process. Information that is more salient is easier to notice, which helps discriminating this relevant information. Gentner and Markman (1997) state that "it is when a pair of items is similar that their differences are likely to be important" (p.51). This is certainly the case in radiological images, where differences between the normal image and the disease image signal pathology. Thus, in order to make the relevant, disease-related information more salient on a radiograph, an image that shows no abnormalities (i.e., a normal image) is the best contrasting image. The normal anatomy on both the normal image and the pathological image can be aligned to each other. The disease-related information, which signifies the main difference between the two images, will then become salient. Saliency influences visual attention and thus makes it easier for students to discriminate disease-related information from irrelevant information.

Additional indications for a positive effect of comparison come from Hammer and colleagues (Hammer, Brechmann, Ohl, Weinshall, & Hochstein, 2010; Hammer, Diesendruck, Weinshall, & Hochstein, 2009). In a neuroimaging study of category-learning, they showed that brain areas associated with directed attention mechanisms become active when participants compare stimuli that do not belong to the same category (Hammer et al., 2010). Their participants studied categories of complex visual stimuli (computer-generated creatures). They did so by either comparing pairs of stimuli from the same category, or pairs of stimuli from different categories. Stimuli could be distinguished from each other based on four different features, such as the color of the eyes. Neural activity was measured with fMRI. Hammer et al. (2010) concluded that the directed attention mechanisms activated by comparison are aimed to highlight specific information that is necessary to discriminate between categories. Comparably, in radiological images, comparison can highlight (make more salient) the information that is necessary to discriminate between diseases: the relevant, disease-related information. Consequently, Hammer et al. (2009) suggest that for learning visual features of diseases, comparison of contrasting radiographs might help to discern relevant information and discriminate it from irrelevant information.

Discriminating relevant from irrelevant information poses a specific difficulty for some visual tasks. In many complex displays, some of the locations contain relevant information; some locations do not contain relevant information and they can be more or less ignored (see, for example, [Balslev et al., 2011](#), pediatricians pay attention specifically to the body part that is moving abnormally, while ignoring the rest of the information). However, the location of relevant information is not always that restricted, which leads to different expertise effects. For example in a task of carefully inspecting abstract paintings, the relevant information is not confined to a specific location that can be attended to: it is globally present over the image. [Zangemeister, Sherman, and Stark \(1995\)](#) found that expert artists looked more globally (with less fixations that are close to each other) at abstract pictures in comparison to novices. This is contradictory to the expertise effect mentioned above, of looking more specifically at relevant locations (fixations close to each other at relevant locations, see [Gegenfurtner et al., 2011](#)).

A comparable phenomenon can be found in diagnosing chest radiographs. Experts are quicker to fixate nodules: small, spherical, often inconspicuous abnormalities ([Nodine & Mello-Thoms, 2010](#)). However, many other types of diseases can be found in the lungs. A distinction can be made between focal and diffuse diseases. Focal diseases lead to lesions at one location, such as a tumor in one lung, while diffuse diseases involve all lobes of both lungs ([Ryu, Olson, Midthun, & Swensen, 2002](#)). Specifically for diffuse diseases, the location of relevant information is not restricted but extends throughout the whole lung (see [Fig. 2](#)).

Consequently, focal diseases require attention that is directed to only one part of the image, the location of the mass or lesion. The disease at this location has to be discriminated from the rest of the image. Diffuse diseases, on the other hand, require attention that is 'directed' to various parts of the image because the disease affects most of the chest. It is therefore very well possible that comparison has a stronger effect on discriminating information indicating focal diseases than on discriminating information indicating diffuse diseases. This is expected to lead to a larger effect of comparison on learning focal diseases, yielding an interaction between type of disease and comparison on visual diagnostic reasoning.

1.2. Hypotheses

The current study investigated whether students learning the radiological appearances of diseases benefit from the comparison of images showing a disease with a normal radiograph. Students

who could compare with normal images during learning were compared with students who could only study images showing diseases. Afterward, students' visual diagnostic skills were tested and students were asked to describe the features of the diseases.

Comparison of radiographs showing abnormalities with 'normal' images is expected to help students discriminate disease-related visual information from the normal anatomy on radiological images ([Gentner & Gunn, 2001](#)), resulting in a higher quality mental representation of the information. The quality of students' mental representation is reflected in their diagnostic reasoning performance as well as the ability to describe the features of diseases, which is a necessary skill for writing radiological reports. However, the focus of the current intervention is on visually discriminating features rather than verbally discriminating features, so the effect of comparison is expected to be more pronounced on a test of diagnostic reasoning than on a feature description test.

Effects of comparison of radiographs showing abnormalities with 'normal' images might be dependent on the type of disease. As explained above, comparison is expected to have a stronger effect on discriminating relevant information indicating focal diseases than on information indicating diffuse diseases, leading to an interaction between type of disease and comparison. So specifically for focal diseases, it was expected that comparison with normal images leads to a higher visual diagnostic accuracy (hypothesis 1) and possibly a higher ability to describe the features of a disease in an image (hypothesis 2).

2. Methods

2.1. Participants

A total of 61 Dutch undergraduate (3rd year) medical students (41 female; 20 male) participated in the study. The mean age of the students was 21.3 years ($SD = 1.15$). Students received a small monetary reward for their participation. They did not have any experience with visual diagnosis in radiology. Students were randomly assigned to the pathology/normal condition ($n = 30$; 22 female, 8 male) and the pathology/pathology condition ($n = 31$; 19 female, 12 male). Test scores of 8 participants (4 students in every condition) were excluded from the analysis due to technical problems during administration of the tests. This resulted in 26 participants in the pathology/normal condition, and 27 participants in the pathology/pathology condition.

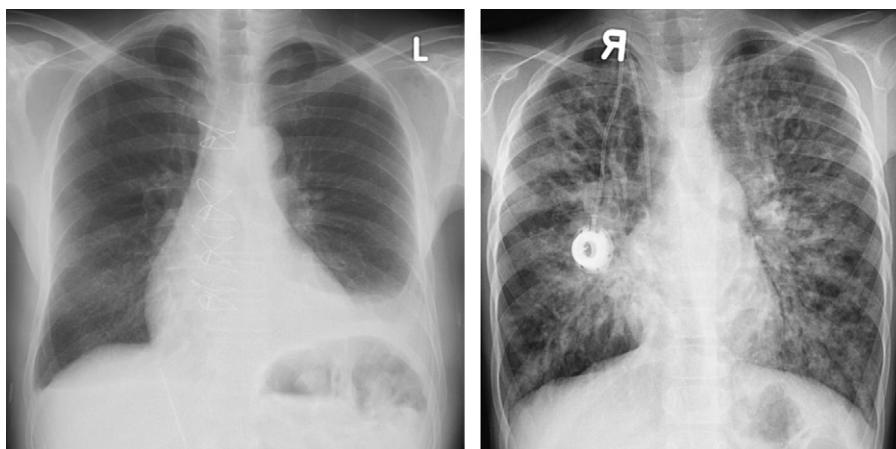


Fig. 2. Example of a focal disease (left: pleural effusion in the left lower lung) and a diffuse disease (right: Cystic Fibrosis).

2.2. Materials

2.2.1. Radiographs

Materials used were 71 Posterior-Anterior conventional chest radiographs of both adults and children. Eight of those radiographs showed no abnormalities. The other 63 radiographs showed in total twelve different diseases of the heart and lungs. The twelve diseases were common diseases that were selected by a senior radiologist. A distinction was made between diseases that led to lesions at one location (focal diseases), and diseases that involved all lobes of both lungs (diffuse diseases) (Ryu et al., 2002). Selected focal diseases were Atelectasis, Cardiomegaly, Lung Tumor, Pleural Effusion, Pneumonia, Pneumothorax, broadened Mediastinum, and enlarged Hila. Selected diffuse diseases were Cystic Fibrosis, Chronic Obstructive Pulmonary Disease (COPD), Lung Metastases, and Miliary TBC. For each of the diseases, at least three good-quality pictures were selected. All personal information was removed from the images. As different diseases can result in the same radiological features, special care was taken to select diseases that can be discriminated from each other based on a chest radiograph. All diseases were part of the curriculum so students were expected to be familiar with all of them, although they had no knowledge whatsoever of what the diseases looked like on a radiograph. All images were selected to show only the indicated disease and were typical examples of that disease.

In the learning phase, each screen showed two radiographs with the name of the disease present in each. In the pathology/normal condition (experimental condition), a radiograph of a patient and a normal image were shown next to each other. In the pathology/pathology condition, two radiographs of patients with the same disease were shown next to each other. A screenshot of the pathology/normal condition is shown in Fig. 3. As described in the Introduction section, it is specifically expected that comparison with a normal image is effective. Therefore, the control condition uses comparison of images showing the same disease, a condition which is not expected to be effective for discrimination. In this way, students in the two conditions are expected to take a similar amount of time for learning and receive exactly the same amount of information. This is preferred over sequential learning, as students in a sequential learning condition often take less time for learning than students in a comparison learning condition (see, for example Ark et al., 2007), making it hard to distinguish between the effect of comparison and a simple time on task effect.

The 12 screens were presented twice, with the diseases in a different order for the first run and the second run. In the

pathology/pathology condition, the same two images of each disease were shown in both the first and second run. In the pathology/normal condition, one image of the disease was shown in the first run, and the other image of the same disease was shown in the second run – always together with a normal image. Normal images were matched to the images of the diseases based on exposure time of the radiograph, age of the patient, and conspicuous but non-relevant anatomical variations such as the presence of breasts or the length of the chest. Matching aimed to make the normal image as similar as possible to the disease image, except for disease-related features.

2.2.2. Visual diagnosis test

The visual diagnosis test consisted of 59 items. Each item consisted of a chest radiograph for which students were required to give a diagnosis by typing it in a textbox. Students were informed that patients shown might have any of the 12 diseases, or might be healthy. Normal images were included to make the task more authentic, but were not incorporated in the test score because students in the pathology/normal condition were exposed to normal images during learning, while students in the pathology/pathology condition were not. Fifty-two of the images showed one of the learned diseases, seven images showed no abnormalities. Disease-items consisted of both images of patients seen during the learning-phase and images of new patients. Five items were deleted because they correlated negatively with the total score. Cronbach's alpha for the visual diagnosis test after deletion of the negatively correlated items was .82. For the first 27 items, students were asked to type the name of the disease that they thought was on the image (one of the 12 learned diseases) or to type "healthy" in case of no abnormalities (thus 13 options in total). For the next 27 items, students were presented with 4 answer options to choose from: the correct answer; "healthy", and two incorrect options. For the four option items, Cronbach's alpha was .68; for the thirteen options items, Cronbach's alpha was .69. Because all items measured the same construct, they were pooled for all statistical analyses to increase power. The visual diagnosis test score was the proportion of correctly diagnosed images. The order of the items was randomized over participants.

2.2.3. Feature description test

The feature description test consisted of 12 radiographs, one for each disease. The same radiographs were presented as in the learning phase. Students were given the correct diagnosis and were required to describe the appearance of each disease on

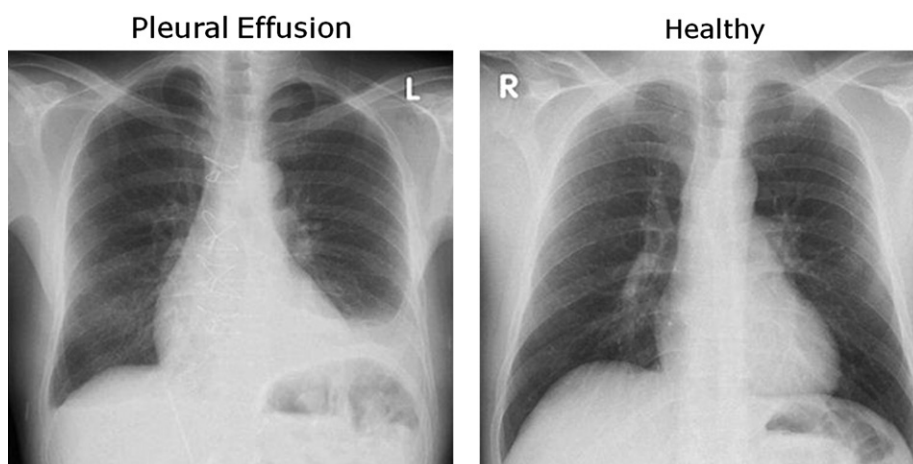


Fig. 3. Screenshot of the learning phase of the pathology/normal condition, showing a patient with pleural effusion on the left and a radiograph without abnormalities on the right.

a conventional radiograph. The feature description test score was the proportion correctly described features of the disease. Features that had to be described were established by an expert radiologist before scoring. Only features that could be described based on the given image were scored. For example, three features could be described for the image of COPD: A small and thin heart, a flat diaphragm, and a blackish appearance of the lungs. Answers of students were scored by one researcher who was blind to the condition of the students. Another researcher scored answers of 30 students to determine inter-rater agreement. The inter-rater agreement was calculated for each disease, using Cohen's Kappa. Kappa per disease ranged from .58 to .92; the mean Kappa was .77, which is considered acceptable. Differences were resolved by discussion between the two scorers, until agreement was reached.

2.3. Procedure

Participants were tested in six experimenter-supervised groups of up to 18 students, but instructed to work individually on their own computer and not to consult with peers. The experiment consisted of a learning phase and a test phase and took approximately 1 h. In the learning phase, participants studied the radiological appearances of the 12 diseases. Diseases were shown one by one on a computer screen, for a maximum of 30 s each. Students could choose to continue to the next disease earlier by hitting the F1 button. After the 12 diseases were shown once, they were all shown again in a different order.

After the learning phase, students took the visual diagnosis test. Students were asked to type in the most likely diagnosis for 59 images while viewing the image. This test was self-paced. Subsequently, students took the feature description test: students were presented with an image of each of the diseases. The image of the disease was accompanied by the name of the disease. Students were required to describe how they could recognize that disease on a conventional radiograph. They could type in their answer in a textbox while viewing the radiograph. This test was also self-paced. Finally, students were thanked for participation.

3. Results

Mixed ANOVAs were conducted with type of disease (focal, diffuse) as within-subjects factor and condition (pathology/normal, pathology/pathology) as between-subjects factor. Significance level was set to $p = .05$. Effect sizes for ANOVA's are reported using η_p^2 , with .01 indicating a small effect, .06 indicating a moderate effect, and .14 indicating a large effect. Effect sizes for separate t -tests are reported using Cohen's d , .2 is considered a small effect, .5 a moderate effect and .8 a large effect.

3.1. Pre-analyses

No significant difference between the two conditions was found on the mean time spent studying the images, $F(1, 51) = 0.05$, $p = .82$, $\eta_p^2 = .001$. The maximum time was 30 s for each screen. The mean learning time per item for focal diseases (pathology/normal condition: $M = 16.4$ s, $SD = 6.8$; pathology/pathology condition: $M = 16.1$ s, $SD = 5.4$) was significantly higher than the mean learning time per item for diffuse diseases (pathology/normal condition: $M = 15.2$ s, $SD = 7.5$; pathology/pathology condition: $M = 14.8$ s, $SD = 5.5$), $F(1, 51) = 16.6$, $p < .001$, $\eta_p^2 = .25$. The interaction between type of disease and condition was not significant, $F(1, 51) = 0.05$, $p = .83$, $\eta_p^2 = .001$. Because there was no significant difference between the conditions on time on task, this variable was not included in subsequent analyses.

The number of false negatives (i.e., reporting "healthy" when a disease was present) on the visual diagnosis test was very low, with a mean of 3.62 ($SD = 2.45$) from 52 items. The number of false positives (reporting a diseases when there was none) was relatively high with a mean of 3.53 ($SD = 1.53$) from 7 items. Three participants did not report any healthy images. It seems that students have a strong bias toward reporting any disease. As in medical school the focus is on diagnosing diseases, this seems in line with the way students are trained.

No significant differences were found on the number of false negatives on the visual diagnosis test, between the pathology/normal condition ($M = 4.04$, $SD = 2.49$) and the pathology/pathology condition ($M = 3.22$, $SD = 2.39$), $t(51) = 1.22$, $p = .23$, Cohen's $d = .34$. Also, the number of false positives did not differ significantly between the pathology/normal condition ($M = 3.27$, $SD = 1.28$) and the pathology/pathology condition ($M = 3.78$, $SD = 1.72$), $t(51) = -1.22$, $p = .23$, Cohen's $d = .34$. Because no significant differences between conditions were found, these variables were not used in subsequent analyses.

3.2. Visual diagnosis test

A significant main effect of type of disease on the visual diagnosis test score was found, $F(1, 51) = 31.27$, $p < .001$, $\eta_p^2 = .38$. The mean score was higher for diffuse diseases ($M = .70$, $SD = .16$) than for focal diseases ($M = .58$; $SD = .17$; see Fig. 4). The main effect of comparison type was not significant, $F(1, 51) = 1.63$, $p = .21$, $\eta_p^2 = .03$; however, a significant interaction effect on the visual diagnosis test score was found, $F(1, 51) = 4.09$, $p = .048$, $\eta_p^2 = .07$, indicating that the effect of comparison with normal was more positive for focal diseases than for diffuse diseases (focal diseases: pathology/normal: $M = .63$, $SD = .15$; pathology/pathology: $M = .54$, $SD = .17$; diffuse diseases: pathology/normal: $M = .70$, $SD = .15$; pathology/pathology: $M = .70$, $SD = .17$). Separate t -tests for focal and diffuse diseases revealed a significant effect of comparison type on focal diseases, $t(51) = 2.08$, $p = .04$, Cohen's $d = .56$. The effect of comparison type on diffuse diseases was not significant, $t(51) = .16$, $p = .87$, Cohen's $d = 0.0$. Note that the significant difference in difficulty between focal and diffuse items (i.e., main effect of type of image) is trivial. It is the consequence of the specific images that were used for this experiment, rather than a property of the types of diseases.

If students made mistakes, they often diagnosed the item as another diffuse disease if the item was a diffuse disease, and

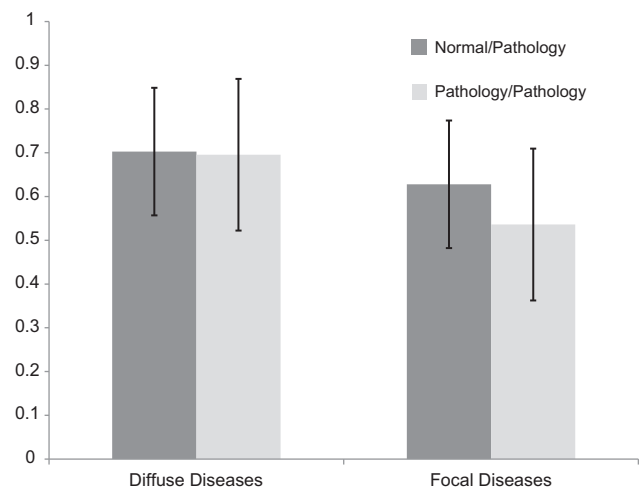


Fig. 4. Mean proportion correct for focal and diffuse diseases. Error bars represent standard deviations.

especially as *another* focal disease if the item was a focal disease. For focal diseases, on average 71.9% of the incorrect responses was a focal disease; for diffuse diseases, on average 47.8% of the incorrect responses was a diffuse disease.

3.3. Feature description test

The feature description test consisted of 12 items. A significant effect of type of disease was found, $F(1, 51) = 8.06, p = .006, \eta_p^2 = .14$. The proportion of correctly described features was higher for diffuse diseases (pathology/normal condition: $M = .39, SD = .12$; pathology/pathology condition: $M = .38, SD = .11$) than for focal diseases (pathology/normal condition: $M = .35, SD = .10$; pathology/pathology condition: $M = .32, SD = .08$). No significant main effect of comparison with normal was found, $F(1, 51) = 0.97, p = .33$. There was no significant interaction effect, $F(1, 51) = 0.35, p = .56$. For each disease, between two and five features could be mentioned. In 17.0% of the cases, none of the present features was mentioned, in 50.5% of the cases, only one of the features was mentioned and in only 32.5% of the cases, more than two features were mentioned. In Table 1 the answers of a typical student can be found. It can be seen that students have many difficulties verbally discriminating between diseases.

4. Discussion

Medical students studied the radiological appearance of lung and heart diseases on chest radiographs. Half of them could compare the radiographs with images showing no abnormalities (normal images), while the other half could only compare radiographs of patients with the same disease. Students who could compare with normal images outperformed students who could not compare with normal images, for focal diseases but not for diffuse diseases (hypothesis 1). There was no significant difference in the learning time between the conditions, so learning time is unlikely to have caused the significant difference in visual diagnosis test scores between the two conditions on focal items. On the feature description test, no significant effects of comparison were found (hypothesis 2), but this might be caused by difficulties that students have in verbalizing discriminative features of diseases.

Comparison with a normal image was proposed to make it easier for students to discriminate relevant from irrelevant information, by making the features of a disease more salient (Gentner & Gunn, 2001; Hammer et al., 2010). Saliency refers to the conspicuity of a – part of a – stimulus in relation to its environment (Itti & Koch, 2000): It makes particular information stand out from the rest of the image. Making a focal disease stand out from the environment (i.e., the chest) is very effective, because it makes it easier to direct attention to the relevant information. However, for diffuse diseases,

the whole chest is involved so the whole image should become more salient. And when everything stands out, it does not stand out any more! Consequently, attention is not directed to a specific location and discrimination of relevant information is not facilitated. This is analogous to highlighting a few words in a text or highlighting all words in a text. If one word is highlighted, it becomes salient. If all words are highlighted, none of them are more salient than others. Accordingly, we found that the positive effect of comparison with the normal image was present for focal diseases but not for diffuse diseases. This indicates that comparison with normal might indeed make the disease-related information more salient, as expected, and therefore easier to discriminate from irrelevant information.

It is not yet clear how attention is directed in diffuse diseases. It seems unlikely that directed attention to a specific relevant location is applicable to diffuse diseases, because there is no location in the image that is more informative than other locations for learning the disease. This is in contrast to focal diseases, where scrutiny of one location yields the information necessary for learning to discriminate the disease. Presumably, another distribution of attention takes place. Using eye-tracking, Nodine, Locher, and Krupinski (1993) describe two patterns of attention in art perception. A focal pattern of attention is characterized by long gazes (clusters of fixations of more than 400 ms) and little coverage of the picture. The goal of this pattern of attention is focal scrutiny of information. A global pattern of attention is characterized by short gazes (clusters of fixation of less than 300 ms) and larger coverage. The goal of this pattern of attention is global surveying and exploration of the picture. Although both patterns of attention might occur simultaneously, the focal pattern of attention may be more pertinent for focal diseases and the global pattern of attention may be more pertinent for diffuse diseases. Clustering of fixations on relevant information – as found in the focal pattern – leads to better understanding of that information (cf. Boucheix & Lowe, 2010). However, for diffuse diseases, the focal pattern of attention seems inadequate and it might be true that the global pattern of attention is more useful for these images. Further research using eye-tracking is necessary to understand attention patterns in diffuse diseases and how these can be influenced to enhance learning.

Incorporating relevant, disease-related information into mental representations requires *finding* which locations hold the relevant information (where is it?), and *interpreting* this information (what makes it what it is?) (Krupinski, 2010). Interpretation is not necessarily verbal and results in a visual representation of the information, which is incorporated in the mental representation of the disease, for use in visual diagnosis. Comparison learning could affect both finding and interpretation of relevant information. Although the setup of the present study did not allow for separate analysis of these two processes, further research should try to disentangle the effects of comparison on finding the location of information and interpreting this information. For example, finding the information could be investigated by presenting images for a very short time span to see whether this is sufficient to detect the disease without the possibility to interpret it. The effect on interpretation alone could be isolated using eye-tracking. It can be checked whether a lesion is fixated on for a sufficient amount of time to be detected, which indicates that any mistakes in diagnosis must be based on interpretation errors (see for example, Manning, Ethell, & Donovan, 2004). Another important issue for further research is whether both comparison with normal and comparison with pathological images are more effective than learning without the opportunity to compare (sequential learning). However, as explained in the Methods section, this requires that time on task can be properly controlled for.

The feature description test showed no significant main effect of comparison and no significant interaction of comparison with

Table 1
Answers of a typical student on all 12 items of the feature description test.

Disease	Description of the student
Lung metastases	"white, round spots"
Cardiomegaly	"enlarged heart, big white area left"
Broadened Mediastinum	"broad, white area from top to bottom in the middle"
Lung Tumor	"white spot"
Pleural Effusion	"white haze"
Atelectasis	"white area in the lungs"
COPD	"dark, black lungs"
Pneumothorax	"white area in the lung on the side"
Enlarged Hila	"large white areas left and right of the middle"
Miliary TBC	"white, speckled lung"
Cystic Fibrosis	"dark lungs with many white stripes"
Pneumonia	"white spot"

Note. Answers are translated from Dutch.

type of disease (hypothesis 2). Inspection of the data of the feature description test shows that students have many difficulties describing the features of diseases. This makes it hard to interpret their scores. On average, only one-third of the features that were present were actually mentioned. The main problem of the students is in verbally discriminating diseases from each other that they are able to discriminate visually, because the learning phase focused on visual rather than verbal discrimination. In half of the cases, only one feature was mentioned, and often, a student only described seeing a white area or white spot in the lungs.

This can be considered a surprising finding, since analysis of verbal data is a common way to investigate learning processes (Fox, Ericsson, & Best, 2011). However, it is well known that some processes are not easy to verbalize (Schooler, 2011). It seemed that the third-year students in our sample did not have the radiological vocabulary that radiologists have, which is necessary to verbally discriminate between features that can be discriminated perceptually. This can be appreciated in Table 1. Consequently, for further research with medical students, non-verbal measures (e.g., eye-tracking data) need to be used in order to understand the processes involved in visual learning.

In conclusion, it was found that comparison with a normal image facilitated learning the appearance of focal diseases but not of diffuse diseases. It seems to be the case that comparison learning was effective mainly because it influenced the saliency of relevant information, which made it easier to discriminate this information and incorporate it in the mental representation of the disease. Comparison learning has hardly been applied to real-life complex visual tasks. We investigated its use in the domain of radiology, but we think it could apply to much more types of complex real-life visual tasks. We expect that comparison with a carefully selected contrasting stimulus is also useful for learning other visual diagnostic tasks in medicine, for example, in dermatology and pathology. Students learning in these domains face comparable problems as students learning radiology and might thus profit from comparison in the same way as radiology students do. For example, it is harder for students compared to pathologists, to discriminate relevant from irrelevant locations in pathology slides (Krupinski et al., 2006). Comparison with slides that show no indications of diseases could help those students. We expect to find an effect of focal versus diffuse diseases in pathology as well, but further research should be done to investigate this. Other complex tasks that require visual skills might also benefit from comparison, such as interpretation of weather maps (e.g., Lowe, 2005), reading of radar screens, and biological classification of birds. For example, in order to learn patterns on weather maps that lead to thunderstorms, students can compare a weather map showing features that will eventually lead to a thunderstorm with a weather map showing a situation that will not end in a thunderstorm. Comparison could help students to see which information is relevant for predicting thunderstorms. Based on our findings, we would predict that comparison would be especially helpful for learning isolated features rather than the patterns that extend across the map, while Lowe hypothesizes that comparison should be especially effective for learning the patterns across the map (Lowe, 2005). Further research should investigate the specific processes that explain how comparison helps learning in complex visual domains, specifically with regard to differences between focal and diffuse patterns.

Visual diagnosis in radiology is a very difficult skill to learn for medical students (Gunderman, Williamson, Fraley, & Steele, 2001). It is highly complex and it takes thousands of cases (Lesgold et al., 1988) before a novice knows 'how to see'. But the pressure on students and residents to learn quickly is high, because there is so much other work to do. Finding new ways to make the learning process more effective and efficient is therefore of utmost

importance and should receive much more attention because, as Gunderman (2012) states: "Education Matters"!

Acknowledgments

The authors would like to thank H.P.A. Boshuizen for giving valuable feedback on the manuscript.

References

- Andrews, J. K., Livingston, K. R., & Kurtz, K. J. (2011). Category learning in the context of co-presented items. *Cognitive Processing*, 12, 161–175. <http://dx.doi.org/10.1007/s10339-010-0377-5>.
- Ark, T. K., Brooks, L. R., & Eva, K. W. (2007). The benefits of flexibility: the pedagogical value of instructions to adopt multifaceted diagnostic reasoning strategies. *Medical Education*, 41, 281–287. <http://dx.doi.org/10.1111/j.1365-2929.2007.02688.x>.
- Balslev, T., Jarodzka, H., Holmqvist, K., de Grave, W., Muijtjens, A. M. M., Eika, B., et al. (2011). Visual expertise in paediatric neurology. *European Journal of Paediatric Neurology*, 16, 161–166. <http://dx.doi.org/10.1016/j.ejpn.2011.07.004>.
- Biederman, I., & Shiffrar, M. (1987). Sexing day-old chicks: a case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 640–645. <http://dx.doi.org/10.1037/0278-7393.13.4.640>.
- Boucheix, J. M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction*, 20, 123–135. <http://dx.doi.org/10.1016/j.learninstruc.2009.02.015>.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137, 316–344. <http://dx.doi.org/10.1037/a0021663>.
- Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. H. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, 22, 47–61. <http://dx.doi.org/10.1016/j.learninstruc.2011.06.002>.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, 23, 1–30. <http://dx.doi.org/10.1007/s10648-011-9174-7>.
- Gentner, D. (2010). Bootstrapping the mind: analogical processes and symbol systems. *Cognitive Science*, 34, 752–775. <http://dx.doi.org/10.1111/j.1551-6709.2010.01114.x>.
- Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory & Cognition*, 29, 565–577. <http://dx.doi.org/10.3758/BF03200458>.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45–56. <http://dx.doi.org/10.1037/0003-066X.52.1.45>.
- Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts. <http://dx.doi.org/10.1126/science.168.3934.958>.
- Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquisition and analogical transfer? *Canadian Journal of Psychology—Revue Canadienne De Psychologie*, 46, 539–550. <http://dx.doi.org/10.1037/h0084333>.
- Gunderman, R. B. (2012). Education matters. *Academic Radiology*, 19, 117–118. <http://dx.doi.org/10.1016/j.acra.2011.10.015>.
- Gunderman, R., Williamson, K., Fraley, R., & Steele, J. (2001). Expertise: implications for radiological education. *Academic Radiology*, 8, 1252. [http://dx.doi.org/10.1016/S1076-6332\(03\)80708-0](http://dx.doi.org/10.1016/S1076-6332(03)80708-0).
- Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison processes in category learning: from theory to behavior. *Brain Research*, 1225, 102–118. <http://dx.doi.org/10.1016/j.brainres.2008.04.079>.
- Hammer, R., Brechmann, A., Ohl, F., Weinshall, D., & Hochstein, S. (2010). Differential category learning processes: the neural basis of comparison-based learning and induction. *Neuroimage*, 52, 699–709. <http://dx.doi.org/10.1016/j.neuroimage.2010.03.080>.
- Hammer, R., Diesendruck, G., Weinshall, D., & Hochstein, S. (2009). The development of category learning strategies: what makes the difference? *Cognition*, 112, 105–119. <http://dx.doi.org/10.1016/j.cognition.2009.03.012>.
- Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: the critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education*, 8, 17–26. <http://dx.doi.org/10.1023/a:1022687404380>.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489–1506. [http://dx.doi.org/10.1016/S0042-6989\(99\)00163-7](http://dx.doi.org/10.1016/S0042-6989(99)00163-7).
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: how experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20, 146–154. <http://dx.doi.org/10.1016/j.learninstruc.2009.02.019>.
- Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, 6, 53–84. <http://dx.doi.org/10.1016/j.plrev.2008.12.001>.
- Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, 72, 1205–1217. <http://dx.doi.org/10.3758/APP.72.5.1205>.

- Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., et al. (2006). Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human Pathology*, 37, 1543–1556. <http://dx.doi.org/10.1016/j.humpath.2006.08.024>.
- Lesgold, A., Robinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: diagnosing x-ray pictures. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.
- Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education*, 14, 225–244. <http://dx.doi.org/10.1007/BF03172967>.
- Lowe, R. K. (2005). Multimedia learning of meteorology. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.
- Manning, D. J., Ethell, S. C., & Donovan, T. (2004). Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph. *British Journal of Radiology*, 77, 231–235. <http://dx.doi.org/10.1259/bjr/28883951>.
- Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. *Psychological Science*, 8, 363–367. <http://dx.doi.org/10.1111/j.1467-9280.1997.tb00426.x>.
- Mettler, F. A. (2005). *Essentials of radiology*. Philadelphia: Elsevier Saunders.
- Nodine, C. F., Locher, P. J., & Krupinski, E. A. (1993). The role of formal art training on perception and aesthetic judgment of art compositions. *Leonardo*, 26, 219–227. <http://dx.doi.org/10.2307/1575815>.
- Nodine, C., & Mello-Thoms, C. (2010). The role of expertise in radiologic image interpretation. In E. Samei, & E. Krupinski (Eds.), *The handbook of medical image perception and techniques* (pp. 139–156). Cambridge: Cambridge University Press.
- Rittle-Johnson, B., & Star, J. R. (2011). The power of comparison in learning and instruction: learning outcomes supported by different types of comparisons. In J. P. Mestre, & B. H. Ross (Eds.), (2011). *Cognition in Education*, Vol. 55 (pp. 199–226). Oxford: Academic Press.
- Ryu, J. H., Olson, E. J., Midthun, D. E., & Swensen, S. J. (2002). Diagnostic approach to the patient with diffuse lung disease. *Mayo Clinic Proceedings*, 77, 1221–1227. <http://dx.doi.org/10.4065/77.11.1221>.
- Samei, E., Flynn, M. J., Peterson, E., & Eyler, W. R. (2003). Subtle lung nodules: influence of local anatomic variations on detection. *Radiology*, 228, 76–84. <http://dx.doi.org/10.1148/radiol.2273020509>.
- Schooler, J. W. (2011). Introspecting in the spirit of William James: comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin*, 137, 345–350. <http://dx.doi.org/10.1037/a0022390>.
- Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychological Science*, 16, 145–151. <http://dx.doi.org/10.1111/j.0956-7976.2005.00795.x>.
- Wood, B. P. (1999). Visual expertise. *Radiology*, 211, 1–3.
- Zangemeister, W. H., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia*, 33, 1009–1025. [http://dx.doi.org/10.1016/0028-3932\(95\)00014-T](http://dx.doi.org/10.1016/0028-3932(95)00014-T).